

How useful are screening instruments for toddlers to predict outcome at age 4? General development, language skills, and symptom severity in children with a false positive screen for autism spectrum disorder.

Mieke Dereu, Herbert Roeyers, Ruth Raymaekers, Mieke Meirsschaut, and Petra Warreyn

Research Group Developmental Disorders, Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium.

Tel: +32(0)92649412

Fax: +32(0)92646489

E-mail: Mieke.Dereu@UGent.be

Abstract

Screening instruments for autism spectrum disorder (ASD) often generate many false positives. It is argued that these children may have other developmental difficulties and are also in need of thorough assessment and early intervention. The current study looked at the predictive validity of positive screens on the Checklist for Early Signs of Developmental Disorders (CESDD) and the Early Screening of Autistic Traits questionnaire (ESAT) at age 2 towards language, cognitive function, and symptom severity at age 4. Children who screened positive on the ESAT scored lower for both language and cognitive functioning at age 4 compared to children who screened negative on the ESAT. Also, the more signs of ASD that were recognized on the CESDD or ESAT, the lower the scores for language and cognitive functioning at age 4. False positive screens could be differentiated from true positive screens on the CESDD only in symptom severity score on the Autism Diagnostic Observation Schedule (ADOS). It seems that early screeners for ASD also detect children with other developmental disorders and that diagnostic instruments such as the ADOS are warranted to differentiate between children with ASD and other developmental problems.

Key words: *Autism Spectrum Disorder, Screening, Toddlers, Language, General Development, Symptom Severity*

How useful are screening instruments for toddlers to predict outcome at age 4? General development, language skills and symptom severity in children with a false positive screen for autism spectrum disorder.

In recent years, many new screening instruments have been developed in an attempt to identify children with an elevated risk for autism spectrum disorder (ASD) as soon as possible. Examples are the Early Screening of Autistic Traits questionnaire (ESAT) [1-2] and the Checklist for Early Signs of Developmental Disorders (CESDD) [3] for children younger than 3 years, the First Year Inventory (FYI) [4] for children of 12 months old, and the Baby and Infant Screen for Children with aUtism Traits (BISCUIT) [5-6] for children of 17 to 37 months of age. In addition, adaptations of existing screening instruments were made to optimize early detection of ASD. For example, the Checklist for Autism in Toddlers (CHAT) [7] for 18 months old children, has an adaptation into a parent questionnaire with the Modified CHAT (M-CHAT) [8] and even more recently the dichotomous nature of the items was revised into a five-point scale with the Quantitative CHAT (Q-CHAT) [9].

This growing number of screening instruments in recent years resulted from the commonly accepted idea that early identification is needed to facilitate early intervention [10]. Several studies have shown that intervention, if started in the preschool years, has benefits for the outcome of children with ASD. It can improve their developmental functioning, but can also lead to decreased maladaptive behaviours and symptom severity [11]. Also, early intervention can be cost effective because secondary behaviour problems may be prevented or treated sooner, lowering the intervention costs during the preschool years [12]. In addition, early detection of ASD facilitates educational planning and makes it possible to monitor and study the early development of children with ASD [13]. Finally, early detection through screening can prevent unnecessary referrals before receiving the appropriate diagnosis and parents can be informed early on about the odds of having another child with ASD [14].

Although the benefits of screening for ASD are clear, no screening instrument for young children has yet been developed with appropriate values for all of the major measures of diagnostic

accuracy: sensitivity (Se; the percentage of children with ASD that screened positive), specificity (Sp; the percentage of children without ASD that screened negative), positive predictive value (PPV; the percentage of children with a positive screen that have ASD), and negative predictive value (NPV; the percentage of children with a negative screen that do not have ASD). So, in screening for ASD, one has to make a choice between a lower number of false positives (i.e., children with a positive screen that do not have the disorder) or a lower amount of false negatives (i.e., children with a negative screen that have the disorder). The former will result in a higher Sp and PPV and the latter in a higher Se and NPV. Many false positive results may unnecessarily worry parents and may adversely lead to expensive and time consuming specialized assessments, whereas many false negatives may result in a delay of diagnosis and consequently a later start of much needed intervention and thereby perhaps a poorer outcome.

Because of the recent emphasis on early detection and intervention for toddlers with ASD, the cut-offs of many screening instruments were chosen to optimize Se, which also resulted in a lower PPV and many false positive screens. However, several studies have stated that the children with a false positive screen for ASD often have other developmental difficulties that also warrant further assessment, correct diagnosis, and treatment. For example, Kleinman and colleagues [15] and Pandey and colleagues [16] reported that the M-CHAT resulted in a considerable false positive rate in children between 16 and 30 months, but in both low-risk and high-risk groups these false positive screens mainly consisted of children with language disorders, general developmental delay, or other disorders. Dereu and colleagues [3] found that there were significantly more children with a developmental disorder (other than ASD) or delay in the group of false positive children than in the group of children with a true negative screen, when screening in a population sample with the CESDD.

Also, some large scale population screens with instruments with lower Se revealed that all false positive children had other developmental problems. In the two stages population screen with the ESAT by Dietz and colleagues [1], the false positive screens (75% of all positive screens) had either an intellectual disability, a language disorder, or problems that would fit other diagnostic categories of the DSM-IV-TR [17] or DC: 0-3 [18]. Also Baron-Cohen and colleagues [19] reported that the two false positive screens for ASD in their population study with the CHAT had a developmental delay. This was however based on a rather limited number of positive

screens: Only 12 children out of 16,000 failed the three key items (protodeclarative pointing, gaze monitoring, and pretend play).

Although these studies suggest that the choice for a higher Se is preferable to enable early intervention, but also because false positives may warrant further assessment to detect other disorders, to our knowledge no study looked into the further development of these children with a false positive screen for ASD. It is still unclear if the general developmental delays and language delays often found in these children persist a couple of years later. The study by Wright and Poulin-Dubois [20] is the only one that looked at the relationship between a screening score and concurrent measures of expressive language acquisition. These authors do acknowledge that future research should also look at the predictive ability of screening measures and compare them. So far, studies have only looked at the diagnostic outcome years after children were screened for ASD [15, 21-22].

The current study wanted to fill this void by looking at the further development of children identified as at high risk for ASD through an early screening project. First of all, by following a sample of children prospectively between the ages of 2 and 4, we wanted to examine the diagnostic status at age 4 of children with a false positive screen for ASD on the CESDD or ESAT at age 2. Second, we also explored the predictive validity of screening scores on the CESDD or ESAT at age 2 towards measures of cognitive functioning, language, and autism symptom severity at age 4. Finally, we explored if false positive screen children could be differentiated from true positive cases by their general development and symptom severity at age 2, their general development, language skills, and symptom severity at age 4, as well as by their change in cognitive development and symptom severity between the ages of 2 and 4.

Method

Participants

The sample consisted of 41 at-risk children (26 boys and 15 girls) out of a larger screening study that were seen for further assessments at the university lab because they showed signs of ASD on a screening instrument ($n = 38$) or had language delay ($n = 3$) [3, 23]. All children were screened within a day-care facility when they were on average 22.35 months old ($SD = 6.56$) and with parent questionnaires when they were on average 25.11 months old ($SD = 6.63$). They

were seen at the university lab for further assessment when they were about 2 years old ($M = 28.66$ months, $SD = 6.21$) and were re-evaluated when they were about 4 years old ($M = 49.40$ months, $SD = 1.71$).

During their participation in this prospective longitudinal study, diagnoses of ASD or other developmental disorders were made by an independent multidisciplinary team according to DSM-IV-TR [17] criteria. A diagnosis of ASD was made in 15 children after referral to a university based autism clinic in Flanders. These autism clinics were informed about the available ADOS scores from the first assessment by a written report, but diagnoses were not necessarily based on these scores. At age 4, this diagnosis was confirmed with the Autism Diagnostic Observation Schedule (ADOS) [24] and the Autism Diagnostic Interview – Revised (ADI-R) [25] in our research centre. All children with a clinical diagnosis of ASD scored above the cut-offs for ASD on the revised ADOS scoring algorithms [26] and came within one point on ADI-R social and communication domains, or met the ADI-R autism cut-off on one domain and came within two points on the other [27]. There were two children who received a working diagnosis of ASD at the Ghent University Hospital autism clinic. For these two children, diagnostic status was still unclear at age 4. Both children scored above the cut-off of ASD on the ADOS, but well below the cut-offs on the ADI-R. Moreover, the parents of one child declined reassessment at the autism clinic when the child was older and for the other child, reassessment at an older age is still warranted. These children were omitted from the analyses whenever children with ASD were compared to children without ASD, but were included in analyses with the total at-risk group.

In addition to children with ASD or a working diagnosis of ASD, eight children received a diagnosis of other developmental disorders after referral: three children had an intellectual disability, four children had a general developmental delay, and one child had a mixed expressive/receptive language disorder. Moreover, six children were considered as developmentally delayed based on their results on the Mullen Scales of Early Learning (MSEL) [28] but were not seen by an independent diagnostic centre: three children showed general developmental delay (Early Learning Composite scores below the 10th percentile, i.e., a composite score for the subscales Visual Reception, Fine Motor, Receptive Language, and Expressive Language that fell below 80 [$M = 100$; $SD = 15$]) and three children showed language delay (scores on the subscales Receptive Language and/or Expressive Language below the 10th percentile) at the first assessment with the MSEL in our research centre. The remaining 10 children showed no developmental delay on the MSEL (scores at or above the 10th percentile on all subscales at the first assessment) and were not diagnosed with any DSM-IV-TR or DC:0-3 diagnosis by the time they were 4 years old.

The children with ASD did not differ from the non-ASD group in gender ratio and socioeconomic status score, although for both groups families were from middle to high socioeconomic status. The groups did not differ in age at each assessment at the university lab, although children with ASD were somewhat older than non-ASD children when the screening instruments were filled out (see Table 1).

[INSERT TABLE 1 ABOUT HERE]

Measures

Screening instruments. Two screening instruments for ASD were used in the current study: the CESDD and the ESAT. The CESDD was developed to be easily filled out by child care workers in day-care centres for children between 3 and 39 months old. The instrument consists of 25 dichotomous items. If children fail two or more items, they screen positive for ASD. Preliminary results based on a population screening of 6,808 children revealed good psychometric properties: the CESDD had a sensitivity of .80 and a specificity of .94 [3]. The ESAT is a parent questionnaire and consists of 14 dichotomous items. Children who fail three or more items on the ESAT screen positive for ASD [1-2]. This instrument was validated in a large population sample of 31,724 children. It had a low sensitivity of .23, but a very high specificity of .99 [29].

Autism symptom severity. The ADOS [24] was administered to assess the severity of the symptom representation for each child. This semi-structured, standardized, observational instrument assesses the social and communicative abilities of individuals with possible ASD. The ADOS consists of four modules intended for use in different age groups and different language levels. In the current study, only Module 1 (for use in children with no speech or use of single words) and Module 2 (for children with phrase speech but no fluent speech yet) were used. A classification of autism or ASD is given if the child has scores higher than established cut-off values. A new scoring algorithm was developed by Gotham and colleagues [26] to improve the discriminant ability of the ADOS, which was replicated by several other studies [30-34]. This revised algorithm also includes items related to repetitive behaviours and restricted interests next to items regarding social affect. Moreover, on the basis of the revised algorithm score, an autism severity score ranging from 1 to 10 can be calculated: scores between 1 and 3 indicating non spectrum severity of symptoms, scores 4 and 5 indicating autism spectrum, and scores of 6 or above correspond to severity of symptoms as seen in autism [35].

Language outcome. To measure both receptive and expressive language abilities, the Reynell Developmental Language Scales – Dutch version (RTOS) was used [36]. Normative data (percentile scores and age equivalents) are available for Dutch speaking children for the subscales Receptive Language, Expressive Language, and for the total score obtained by the child.

General developmental level. This was measured with the MSEL [28], an instrument that assesses the cognitive functioning of children between 0 and 68 months. Normative scores are provided for the five subscales (Gross Motor, Visual Reception, Fine Motor, Receptive Language, and Expressive language) and the MSEL also yields a standard score representing general development, cognitive functioning, or intelligence: the Early Learning Composite score (ELC; $M = 100$, $SD = 15$; Mullen, 1995). This score was used in the current study to represent the developmental quotient (DQ) of the children.

Early intervention. All parents completed a survey after the last assessment in which they gave information about the education their child received and whether or not their child received any treatment. If children received treatment, parents gave information about the nature of the treatment (speech therapy, motor therapy, psychotherapy, occupational therapy, or other [e.g., group therapy or parental guidance]), the duration, and intensity of the treatment (when did each type of treatment started and [if applicable] ended, how many minutes lasted each session, and what was the frequency of these sessions). Regarding education, some children were enrolled in a special education program. Also, some children who were enrolled in a regular school, received some additional individual assistance for a couple of hours per week. Table 2 gives an overview of the treatment and education children received during their participation to the study. Overall, only five out of the 26 children without ASD received some treatment or special education, whereas 14 out of the 15 children with ASD received treatment, additional assistance at school, or were enrolled in a special education program during their participation in the current study.

[INSERT TABLE 2 ABOUT HERE]

Procedure

As part of a large scale screening study in Flemish day-care facilities, 7,092 children were screened for ASD with the CESDD by child care workers between February 2006 and April 2010. Parents were asked to fill out additional questionnaires when their child screened positive for ASD on the CESDD or failed a language milestone. Some parents also took part in an unpublished validation study on the Dutch translation of the M-CHAT. All children who exhibited signs of ASD or language delay on one of the parent questionnaires used, were invited for further

evaluation at the university lab with the ADOS and MSEL (for more details on the screening procedure, see Figure 1) [3, 23]. Only the ESAT was filled out for all children and is therefore the only parent questionnaire reported in this study. Based on the results of the ADOS, MSEL, and the clinical judgment of the research team, children were referred for a diagnostic evaluation if necessary.

A total of 79 children were seen at the university lab. This sample consisted of the 70 children seen at the university lab who have previously been described by Dereu and colleagues [23]. These 70 children were selected based on their CESDD results. An additional nine children with a negative screen and no language delay on the CESDD were invited for further assessment because they screened positive on the M-CHAT (see Figure 1). All children younger than 42 months at initial assessment who turned 4 before October 2010 (when the study was finished), were invited for a re-evaluation around their fourth birthday with the ADOS and MSEL. In addition, the RTOS was administered as a language outcome measure and the ADI-R was administered whenever children received a diagnosis of ASD or a working diagnosis of ASD after referral. The parents of 41 out of 72 children that met the inclusion criteria consented with re-evaluation around age 4, resulting in a compliance rate of 56.94%. Compliant and non-compliant children did not differ in gender and screening results on the CESDD and ESAT, $\chi^2(1)$ ranged from 0.66 to 2.71, all $p > .05$. Also, compliant and non-compliant children did not differ in SES, age at screening and first assessment, developmental quotient and symptom severity at initial assessment, $F(1,70)$ ranged from 0.05 to 2.00, all $p > .05$.

The first assessment took place as soon as possible after the screening instruments were filled out for the children, on average 3.55 months ($SD = 4.83$) after parent questionnaires were filled out and 6.31 months ($SD = 5.49$) after the screening within the day-care centres. Re-evaluation took place about 20.73 months ($SD = 5.37$) after the initial assessment. Written informed consent was obtained from all parents prior to the first assessment at the university lab.

[INSERT FIGURE 1 ABOUT HERE]

Results

Description of diagnostic outcome of children with a positive screen for ASD

Of the 31 children with a positive screen for ASD on the CESDD, 11 children were diagnosed with ASD, two children received a working diagnosis of ASD, 14 children had other developmental problems, and only four children received no diagnosis yet. Of the 12 children with a positive screen on the ESAT, six children were diagnosed with ASD, five children had other developmental problems, and only one child received no diagnosis yet. All children with a positive screen on the ESAT screened also positive on the CESDD. Table 3 gives an overview of the screening results on the ESAT and CESDD at age 2 and diagnostic outcome at age 4.

In addition to these children with a positive screen on the CESDD or ESAT, the current study also incorporated 10 children who screened negative on both instruments. Seven children screened positive on another available parent screening instrument (i.e., positive screen for ASD on the FYI, the M-CHAT, and/or the Social Communication Questionnaire [SCQ]) [37]. Only three children screened negative for ASD on all the available screening instruments for the child and were seen for further assessment solely on the basis of language delay and/or other parental concerns. However, two out of these three children received a diagnosis of ASD after referral.

[INSERT TABLE 3 ABOUT HERE]

Differences in outcome variables at age 4 for children with a positive versus a negative screen for ASD at age 2

To assess if children with a positive screen for ASD differed from children with a negative screen in autism severity, language, and cognitive outcome at age 4, MANOVA's were carried out with the outcome variables (age equivalents on the expressive and receptive scales of the RTOS, DQ, and ADOS severity score) as dependent variables and the screening results as factors: one model was tested for the CESDD, one for the ESAT.

Screening positive or not on the CESDD did not lead to differences in outcome at age 4, $F(4,36) = 1.59$, $p = .197$, $\eta^2 = .15$. However, screening positive on the ESAT or not could

differentiate children on the outcome measures, $F(4,36) = 4.27, p = .006, \eta^2 = .32$. Children for whom parents indicated signs of ASD at age 2 had significantly lower scores for receptive and expressive language on the RTOS at age 4 than children who screened negative on the ESAT, $F(1,39) = 16.84, p < .001, \eta^2 = .30$, and $F(1,39) = 9.31, p = .004, \eta^2 = .19$, respectively. Children who screened positive on the ESAT had also a significantly lower DQ than children with a negative screen on the ESAT, $F(1,39) = 17.06, p < .001, \eta^2 = .30$. However, children with a positive or negative screen on the ESAT did not differ in their ADOS severity scores at age 4, $F(1,39) = 1.64, p = .209, \eta^2 = .04$.

To evaluate if the total scores on the CESDD and ESAT were related to outcome measures at age 4, correlation coefficients were calculated. The nonparametric Spearman rho coefficient was chosen because of the skew distribution of the scores on the screening instruments. Table 4 displays the correlation coefficients.

Higher total scores on the CESDD or ESAT around age 2 resulted in (marginally) significantly lower DQ and lower age equivalents for receptive and expressive language at age 4. However, no significant correlations were found between total scores on the CESDD or ESAT and the ADOS severity score at age 4.

[INSERT TABLE 4 ABOUT HERE]

Differentiating false positive screen cases from children with a true positive screen

Children who screened false positive on the CESDD were compared to true positives using a series of multivariate analyses of variance with group (false positive versus true positive) as between-subjects factor. Separate models were tested with as dependent variables first of all measurements at age 2, secondary measurements at age 4 and finally the evolution in cognitive development and symptom severity between the ages of 2 and 4. Growth in cognitive development and symptom severity was computed by a difference score between scores at age 4 and scores at age 2 for respectively the DQs and the ADOS severity scores. Exploratory correlations showed a significant relationship of both the length of the time interval between the assessments and the total amount of early intervention received in hours during participation in this study with the

change in cognitive development and symptom severity. Also, there were no differences found between true and false positives in the amount of early intervention or time interval between assessments. Therefore these variables were entered in the final model as covariates.

Children who screened false positive on the CESDD already differed from children with a true positive screen on the CESDD at age 2, $F(2,24) = 11.43, p < .001, \eta^2 = .49$. This difference was mainly due to differences in ADOS severity scores, $F(1,25) = 21.85, p < .001, \eta^2 = .47$. The DQ of children with a false positive screen did not differ significantly from that of children with a true positive screen at age 2, $F(1,25) = 2.87, p = .103, \eta^2 = .10$. Children with false positive and true positive screens on the CESDD differed also significantly in outcome measures at age 4, $F(4,24) = 27.15, p < .001, \eta^2 = .82$. Children with a false positive screen on the CESDD had significantly lower ADOS severity scores at age 4, $F(1,27) = 102.13, p < .001, \eta^2 = .79$. Also, there was a trend towards a higher DQ at age 4 in children with a false positive screen on the CESDD, $F(1,27) = 3.81, p = .061, \eta^2 = .12$. However, false positive screens did not differ from true positive screens on the CESDD in receptive language at age 4, $F(1,27) = 2.24, p = .146, \eta^2 = .08$, and in expressive language at age 4, $F(1,27) = 2.66, p = .114, \eta^2 = .09$.

Children with a false positive screen and a true positive screen on the CESDD showed also a different trajectory in cognitive development and symptom severity, when controlling for the length of the time interval between the initial and outcome assessment, $F(2,23) = 5.91, p = .008, \eta^2 = .34$. Children with a true positive screen differed significantly from children with a false positive screen in their ADOS severity difference score, $F(1,24) = 6.31, p = .019, \eta^2 = .21$. Children with a true positive screen showed an increase in their severity score, whereas children with a false positive screen had a slight decrease in their ADOS severity score between ages 2 and 4. A trend towards a larger increase in DQ in children with a false positive screen compared to true positive screen children was also noted, $F(1,24) = 3.97, p = .058, \eta^2 = .14$. The mean scores of children with a false positive or true positive screen on the CESDD are shown in Table 5.

Similar analyses of variance could not be performed for screening results on the ESAT, because the amount of children with a true positive screen or a false positive screen was too low. However, for clarity, the mean scores of children with a positive screen on the ESAT are shown in Table 6.

[INSERT TABLE 5 AND 6 ABOUT HERE]

Discussion

The purpose of this study was to determine whether child care workers' and parental reports on early signs of ASD are related to symptom severity, language and general development of the children later on. To our knowledge, the current study is the first one to look at the predictive validity of screening instruments for ASD beyond diagnostic outcome and thereby to examine if developmental problems in positive screen children persist throughout early childhood. For example, although Pandey and colleagues [16] report group differences in outcome measures such as the MSEL and ADOS between two different age groups within a high-risk and low-risk sample, they do not report the predictive validity of the M-CHAT they administered towards these outcome measures. We wanted to fill this void by studying the predictive validity of screening results on the CESDD (child care workers' report) and ESAT (parental report) of 2-year olds towards language abilities, general development and symptom severity at age 4.

Screening positive or negative on the CESDD did not lead to differences in outcome measures at age 4. However, screening positive or negative on the ESAT was related to outcome differences at age 4. Children with a positive screen on the ESAT had a lower DQ and lower age equivalents for both receptive and expressive language than children with a negative screen. However, positive and negative screen children on the ESAT did not differ in their ASD symptom severity at age 4. One should note that within the screening procedure, the CESDD was used as a level one screener and the ESAT as a level two screener. Only parents of children with an elevated risk for ASD based on a positive screen on the CESDD or suspected language delay were asked to fill out the ESAT (see Figure 1). For the current sample this resulted in the fact that all children with a positive screen on the ESAT also screened positive on the CESDD. So one can conclude that if both child care workers and parents report early signs of ASD in the first years of life, these children will probably still have some developmental problems at age 4. Looking at the amount of early signs of ASD reported by child care workers or parents, both total scores on the CESDD and ESAT were significantly related to language and general development at age 4 but not to symptom severity. The results of this limited sample suggest that screening instruments do have a good

predictive validity towards developmental outcome a couple of years later, but not necessarily towards diagnostic outcome in terms of having the disorder they intend to screen for.

Yirmiya and Charman [40] concluded based on a review of both prospective and retrospective studies on early signs of ASD that we do not yet have sufficient insight in which of the early signs of ASD known to us are specific for ASD. In order to address this specificity question, they argued that we need data on children who do and do not continue to develop ASDs later on, after showing these early markers. The current study which reports on the further development of children who showed early signs of ASD (i.e., screened positive for ASD) and who did and did not have ASD at age 4 can be seen as an attempt to address this question. The lack of correlation between total scores on the CESDD and ESAT at age 2 (which reflects the amount of early signs recognized for a child) and the symptom severity scores of the ADOS at age 4 (which is a diagnostic instruments that confirmed all clinical diagnoses of ASD in the current sample) suggest that the early signs included in these screening instruments indeed lack specificity for ASD.

An important issue that arises from this finding is how children who do have the disorder can be differentiated from children with a positive screen who may have other developmental concerns, but do not have ASD. Because of the limited number of children with a positive screen on the ESAT, we only statistically compared children with a true positive screen and children with a false positive screen on the CESDD in trying to answer this question. Results showed that both at 2 years and 4 years false positive and true positive screens differed in their ADOS severity scores with true positive screens having higher severity scores on both assessments. So, well established diagnostic instruments like the ADOS, as part of the clinical assessment following a screening procedure, can distinguish true from false positive screens for ASD. Also, true positive screens and false positive screens followed a different developmental trajectory between 2 and 4 years of age. Although only marginally significant and with limited effect size, children with a false positive screen showed a greater increase in their DQ with most (but not all) of the children scoring within the normal range by the time they are 4 years old. In addition, a significantly different trajectory was noted between children with a true positive and false positive screen on the CESDD in their symptom severity scores on the ADOS. Whereas children with a false positive screen showed a slight decrease in their symptom severity scores, most children with a true positive screen had an

increase in their severity score. Thus, in children who have the disorder, the symptoms may become more apparent between the ages of 2 and 4 [39].

Limitations

A notable limitation in the current sample is the small sample size, which makes it hard to draw conclusions about the predictive ability of the screening instruments in the larger population. Children who were not seen for a re-evaluation at age 4 did not differ significantly from the group of children who took part in the re-evaluation in gender, SES, age at screening and first assessment, screening results on the CESDD and ESAT, and developmental quotient or symptom severity at initial assessment. However, it remains possible that the parents of children who exhibited more severe developmental problems at age 2 were more inclined to react to our invitation for further assessment at the university lab after screening, which was also found in previous population screening studies [23,40]. This may have resulted in more developmentally challenged children with a false positive screen on the CESDD in the current sample compared to the general population. Also, because of the small amount of positive screens on the ESAT, no meaningful comparison between false positive and true positive screens could be made. Finally, there was a low amount of children with a clinical diagnosis of ASD. Therefore, data on Se, Sp, or PPV of the screening instruments could not be estimated. However, this was not our intention, since comparison of Se, Sp, PPV, and NPV of the screening instruments used in the larger population cohort (of which the current sample was drawn) is already provided in a recent publication [23]. Also, these limited amount of clinical diagnosis of ASD were made by independent diagnostic centres. These centres were always informed about the ADOS results of our first assessment, so we cannot rule out that some of these diagnoses were partially based on these ADOS results. Therefore, there is a potential circularity between assigning children to true or false positive screen groups and the difference score of symptom severity we calculated, which was based on the first administration of the ADOS (of which the independent autism clinics received a detailed written report).

Another important limitation is the limited information on early intervention that we collected for the children. The parents gave us some specifics about the interventions their child received during participation in this study. However, there is a large diversity in the nature and

duration of treatment for children with developmental disorders in Flanders, as well as in the education they received. This led to our decision not to incorporate a quantification representing the intensity of treatment received between ages 2 and 4 as a covariate in the analyses. This means that we cannot rule out the effect of interventions on the outcome or the developmental trajectory the children followed. It is however noteworthy that true positive children had a smaller increase in DQ compared to false positive children and an increasing severity score during their participation, while all but one of these children received treatment or special education. False positive screen children had a better prognosis regarding DQ and severity score, although only about one in three children within this group received treatment or special education.

Finally, for some assessments the age range of the children was fairly large. Consequently, the predictive power of the screeners towards outcome at age 4 may be influenced by the time interval between the report of early signs of ASD by parents or child care workers and the final assessment. In this light it is also noteworthy that children were on average a couple of months older when the ESAT was filled out. This could partially explain why correlations found between screening results and outcome measures were somewhat higher for the ESAT than for the CESDD and why screening positive on the ESAT could predict outcome at age 4 whereas screening positive on the CESDD alone could not.

Conclusion

These study results support the idea that many false positive screens for ASD have other developmental problems. Moreover, at least some of these children persist to have developmental difficulties until they are 4 years old. Children with a positive screen for ASD in toddlerhood are therefore an at-risk group that should be monitored during the following years. Because of the low specificity of screeners for ASD, well established diagnostic instruments should be used in further clinical assessments to distinguish true from false positive screen cases.

Acknowledgments

Partial funding for this research was provided by support from Steunpunt Expertisenetwerken and Vlaamse Vereniging Autisme. We thank the day-care centres, the children and their families for their participation to the study. Special thanks goes to the participating diagnostic centra across Flanders (COS and RCA Gent, COS and RCA Antwerpen, COS and RCA Brussel, COS and ECA Leuven).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical standards

This study was approved by the ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University, where the study was conducted.

References

1. Dietz C, Swinkels S, van Daalen E, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum disorder in children aged 14 to 15 months. II: Population screening with the Early Screening of Autistic Traits questionnaire (ESAT). Design and general findings. *J Autism Dev Disord* 36:713-722. doi:10.1007/s10803-006-0114-4
2. Swinkels SHN, Dietz C, van Daalen E, Kerkhof IHGM, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum in children aged 14 to 15 months. I: The development of the Early Screening of Autistic Traits questionnaire (ESAT). *J Autism Dev Disord* 36:723-732. doi:10.1007/s10803-006-0115-0
3. Dereu M, Warreyn P, Raymaekers R, Meirsschaut M, Pattyn G, Schietecatte I, Roeyers H (2010) Screening for autism spectrum disorders in Flemish day-care centres with the Checklist for Early Signs of Developmental Disorders. *J Autism Dev Disord* 40:1247-1258. doi:10.1007/s10803-010-0984-0
4. Reznick JS, Baranek GT, Reavis S, Watson LR, Crais ER (2007) A parent-report instrument for identifying one-year olds at risk for an eventual diagnosis of autism: The First Year Inventory. *J Autism Dev Disord* 37:1691-1710. doi:10.1007/s10803-006-0303-y

5. Matson JL, Wilkins J, Sevin JA, Knight C, Boisjorli JA, Sharp B (2009) Reliability and item content of the Baby and Infant Screen for Children with aUtism Traits (BISCUIT): Parts 1-3. *Res Autism Spectr Disord* 3:336-344. doi:10.1016/j.rasd.2008.08.001
6. Matson JL, Wilkins J, Sharp B, Knight C, Sevin JA, Boisjoli, JA (2009) Sensitivity and specificity of the Baby and Infant Screen for Children with aUtism Traits (BISCUIT): Validity and cutoff scores for autism and PDD-NOS in toddlers. *Res Autism Spectr Disord* 3:924-930. doi:10.1016/j.rasd.2009.04.001
7. Baron-Cohen S, Allen J, Gillberg C (1992) Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *Br J Psychiatry* 161:839-843. doi:10.1192/bjp.161.6.839
8. Robins DL, Fein D, Barton ML, Green JA (2001) The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord* 31:131-144. doi:10.1023/A:1010738829569
9. Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco G, Brayne C (2008) The Q-CHAT (Quantitative Checklist for Autism in Toddlers): A normally distributed quantitative measure of autistic traits at 18-24 months of age: Preliminary report. *J Autism Dev Disord* 38:1414-1425. doi:10.1007/s10803-007-0509-7
10. Filipek PA, Accardo PJ, Banarek GT, Cook EH, Dawson G, Gordon B et al (1999) The screening and diagnosis of autism spectrum disorders. *J Autism Dev Disord* 29:439-484. doi:10.1023/A:1021943802493
11. Rogers SJ, Vismara LA (2008) Evidence-based comprehensive treatments for early autism. *J Clin Child Adolesc Psychol* 37:8-38. doi:10.1080/15374410701817808
12. Jacobson JW, Mulick JA, Green G (1998) Cost-benefit estimates for early intensive behavioral intervention for young children with autism – general model and single state case. *Behav Interv* 13:201-226. doi:10.1002/(SICI)1099-078X(199811)13:4<201::AID-BIN17>3.0.CO;2-R
13. National Research Council (2001) Educating children with autism. National Academy Press, Washington DC
14. Renty J, Roeyers H (2006) Satisfaction with formal support and education for children with autism spectrum disorder: The voices of the parents. *Child: Care, Health & Dev* 32:371-385. doi:10.1111/j.1365-2214.2006.00584.x
15. Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL et al (2008) The Modified Checklist for Autism in Toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *J Autism Dev Disord* 38:827-839. doi:10.1007/s10803-007-0450-9
16. Pandey J, Verbalis A, Robins DL, Boorstein H, Klin A, Babitz T et al (2008). Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism* 12:513-535. doi:10.1177/1362361308094503
17. American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders (4th ed, text rev). American Psychiatric Association, Washington DC
18. ZERO TO THREE: National Center for Infants, Toddlers and Families (1994) Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood. ZERO TO THREE, Washington DC

19. Baron-Cohen S, Cox A, Baird G, Swettenham J, Nightingale N, Morgan K et al (1996) Psychological markers in the detection of autism in infancy in a large population. *Br J Psychiatry* 168:158-163. doi:10.1192/bjp.168.2.158
20. Wright K, Poulin-Dubois D (2011) Modified Checklist for Autism in Toddlers (M-CHAT) screening at 18 months of age predicts concurrent understanding of desires, word learning and expressive vocabulary. *Res Autism Spectr Disord*. doi: 10.1016/j.rasd.2011.04.004
21. Baird G, Charman T, Baron-Cohen S, Cox A, Swettenham J, Wheelwright S, Drew A (2000) A screening instrument for autism at 18 months of age: A 6-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry* 39:694-702. doi:10.1097//00004583-2000006000-00007
22. Eaves LC, Ho HH (2004) The very early identification of autism: Outcome to age 4½-5. *J Autism Dev Disord* 34:367-378. doi:10.1023/B:JADD.0000037414.33270.a8
23. Dereu M, Raymaekers R, Warreyn P, Schietecatte I, Meirsschaut M, Roeyers H (2011) Can child care workers contribute to the early detection of autism spectrum disorders? A comparison between screening instruments with child care workers versus parents as informants. *J Autism Dev Disord*. doi:10.1007/s10803-011-1307-9.
24. Lord C, Rutter M, DiLavore P, Risi S (1999) *Autism Diagnostic Observation Schedule: Manual*. Western Psychological Services, Los Angeles
25. Rutter M, LeCouteur A, Lord C (2003) *Autism Diagnostic Interview Revised: Manual*. Western Psychological Services, Los Angeles
26. Gotham K, Risi S, Pickles A, Lord C (2007) The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *J Autism Dev Disord* 37:613-627. doi:10.1007/s10803-006-0280-1
27. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P et al (2006) Combining information from multiple sources in the diagnosis of autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry* 45:1094–1103. doi: 10.1097/01.chi.0000227880.42780.0e
28. Mullen EM (1995) *Mullen Scales of Early Learning*. American Guidance Service, Circle Pines MC
29. Groen WB, Swinkels SH, van der Gaag RJ, Buitelaar JK (2007) Finding effective screening instruments for autism using Bayes theorem. *Arch Pediatr Adolesc Med* 161:415-416. doi:10.1001/archpedi.161.4.415
30. de Bildt A, Oosterling IJ, van Lang NDJ, Sytema S, Minderaa RB, van Engeland H et al (2011) Standardized ADOS scores: Measuring severity of autism spectrum disorders in a Dutch sample. *J Autism Dev Disord* 41:311-319. doi:10.1007/s10803-010-1057-0
31. Gotham K, Risi S, Dawson G, Tager-Flusberg H, Joseph R, Carter A et al (2008) A replication of the Autism Diagnostic Observation Schedule (ADOS) Revised Algorithms. *J Am Acad Child Adolesc Psychiatry* 47:642-651. doi:10.1097/CHI.0b013e31816bffb7
32. Gray KM, Tonge BJ, Sweeney DJ (2008) Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule with young children with developmental delay: Evaluating diagnostic validity. *J Autism Dev Disord* 38:657-667. doi:10.1007/s10803-007-0432-y

33. Oosterling I, Roos S, de Bildt A, Rommelse N, de Jonge M, Visser J et al (2010) Improved diagnostic validity of the ADOS revised algorithms: A replication study in an independent sample. *J Autism Dev Disord* 40:689-703. doi:10.1007/s10803-009-0915-0
34. Overton T, Fielding C, de Alba RG (2008) Exploratory analysis of the ADOS revised algorithm: Specificity and predictive value with Hispanic children referred for autism spectrum disorders. *J Autism Dev Disord* 38:1166-1169. doi:10.1007/s10803-007-0488-8
35. Gotham K, Pickles A, Lord C (2009) Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord* 39:693-705. doi:10.1007/s10803-008-0674-3
36. Schaerlaekens A, Zink I, van Ommeslaeghe K (2003) Reynell Taalontwikkelingsschalen. Handleiding – Tweede versie. Swets & Zeitlinger, Lisse, The Netherlands.
37. Rutter M, Bailey A, Lord C (2003) Social Communication Questionnaire (SCQ). Western Psychological Services, Los Angeles.
38. Yirmiya N, Charman T (2010) The prodrome of autism: Early behavioural and biological signs, regression, peri- and post-natal development and genetics. *J Child Psychol Psychiatry* 51:432-458. doi:10.1111/j.1469-7610.2010.02214.x
39. Ozonoff S, Iosif AM, Baguio F, Cook IC, Hill MM, Hutman T et al (2010) A prospective study of the emergence of early behavioural signs of autism. *J Am Acad Child Adolesc Psychiatry* 49:256-266. doi:10.1097/00004583-201003000-00009
40. Dietz C, Swinkels SHN, van Daalen E, van Engeland H, Buitelaar JK (2007) Parental compliance after screening social development in toddlers. *Arch Ped Adolesc Med* 161:363-368. doi:10.1001/archpedi.161.4.363

Tables

Table 1

Demographic characteristics of the ASD and non-ASD group

| | ASD group (<i>n</i> = 15) | | | Non-ASD group (<i>n</i> = 24) | | |
|--------------------|-------------------------------|---------------|---------------|-----------------------------------|---------------|---------------|
| | <i>M</i> | (<i>SD</i>) | range | <i>M</i> | (<i>SD</i>) | range |
| SES | 46.33 | (12.91) | 20 – 66 | 50.75 | (11.31) | 12 – 58 |
| Age | | | | | | |
| CESDD | 24.82 | (6.82) | 15.63 – 38.40 | 20.92 | (5.55) | 9.33 – 31.07 |
| ESAT | 27.94 | (6.91) | 17.27 – 41.33 | 23.32 | (5.90) | 11.83 – 33.70 |
| Initial assessment | 30.40 | (7.22) | 20.23 – 43.43 | 27.71 | (5.38) | 17.23 – 37.10 |
| Outcome assessment | 49.58 | (1.97) | 46.37 – 52.93 | 49.26 | (1.60) | 46.50 – 53.63 |
| Male/Female | 10/5 | | | 14/10 | | |
| | | | | $\chi^2(1) = 0.27$ | | |

Note. ASD = autism spectrum disorder, SES = socio-economic status (Hollingshead, 1975), CESDD = Checklist for Early Signs of Developmental Disorders, ESAT = Early Screening of Autistic Traits questionnaire.

[†] $p < .10$. * $p < .05$.

Table 2

Education and treatment characteristics

| | ASD (<i>n</i> = 15) | Non-ASD (<i>n</i> = 24) ^a |
|---|-------------------------|--|
| Education at age 4 | | |
| - Regular education | 1 | 20 |
| - Regular education with additional individual assistance | 9 | 2 |
| - Special education | 5 | 2 |
| Hours/week intervention received prior to age 4 (when applicable) | (<i>n</i> = 8) | (<i>n</i> = 5) |
| - Mean | 2.25 | 2.75 |
| - <i>SD</i> | 0.89 | 1.09 |
| - Range | 1 - 4 | 1.5 - 4 |

^a The two children with a working diagnosis of ASD were not included in this table. Both children attended a regular school and did not receive any treatment.

Table 3

Overview of screening results on the CESDD and ESAT^a

| | | | ESAT | | Total |
|--------|-------|----------|----------|----------|-------|
| | | | Negative | Positive | |
| no ASD | CESDD | Negative | 6 | 0 | 6 |
| | | Positive | 12 | 6 | 18 |
| | | Total | 20 | 6 | 24 |
| ASD | CESDD | Negative | 4 | 0 | 4 |
| | | Positive | 5 | 6 | 11 |
| | | Total | 9 | 6 | 15 |

Note. ASD = autism spectrum disorder, CESDD = Checklist for Early Signs of Developmental Disorders, ESAT = Early Screening of Autistic Traits questionnaire.

^a The screening results of the two children with a working diagnosis of ASD at age 2 and unclear diagnostic outcome at age 4 were omitted from this table. They both screened positive for ASD on the CESDD and negative on the ESAT.

Table 4

Correlations between total scores on screening instruments and outcome measures (N = 41)

| | Early Learning Composite | Symptom severity score | Receptive language AE | Expressive language AE |
|-------|-----------------------------|---------------------------|--------------------------|---------------------------|
| CESDD | -.31 | .23 | -.31 | -.33* |
| ESAT | -.45** | .25 | -.48** | -.41** |

Note. CESDD = Checklist for Early Signs of Developmental Disorders, ESAT = Early Screening of Autistic Traits questionnaire, AE = age equivalent.

* $p < .05$. ** $p < .01$.

Table 5

Comparison of mean scores of children with a true positive versus a false positive screen on the Checklist for Early Signs of Developmental Disorders (CESDD)

| | True positive screens | | | | False positive screens | | | |
|-------------------|-----------------------|---------------|----------|----------|------------------------|---------------|----------|----------|
| | <i>M</i> | (<i>SD</i>) | range | <i>n</i> | <i>M</i> | (<i>SD</i>) | range | <i>n</i> |
| Age 2 | | | | | | | | |
| DQ | 67.73 | (20.11) | 49 – 108 | 11 | 80.38 | (18.36) | 52 – 126 | 16 |
| ADOS severity | 4.91 | (2.12) | 1 – 8 | 11 | 1.88 | (1.26) | 1 – 4 | 16 |
| Age 4 | | | | | | | | |
| RL | 36.55 | (12.43) | 24 – 60 | 11 | 43.00 | (10.54) | 24 – 57 | 18 |
| EL | 35.18 | (10.31) | 24 – 51 | 11 | 42.67 | (12.87) | 24 – 60 | 18 |
| DQ | 82.55 | (36.05) | 49 – 136 | 11 | 105.28 | (26.57) | 51 – 134 | 18 |
| ADOS severity | 6.36 | (1.36) | 5 – 9 | 11 | 1.83 | (1.04) | 1 – 5 | 18 |
| Difference scores | | | | | | | | |
| DQ | 14.82 | (23.47) | -7 – 60 | 11 | 30.31 | (12.89) | -5 – 48 | 16 |
| ADOS severity | 1.45 | (2.58) | -1 – 7 | 11 | -0.31 | (1.40) | -3 – 2 | 16 |

Note. DQ = developmental quotient, ADOS severity = symptom severity score measured with the Autism Diagnostic Observation Schedule module 1 or 2, RL = receptive language age equivalent, EL = expressive language age equivalent.

Table 6

Comparison of mean scores of children with a true positive versus a false positive screen on the Checklist for Early Signs of Developmental Disorders (ESAT)

| | True positive screens | | | | False positive screens | | | |
|-------------------|-----------------------|---------------|----------|----------|------------------------|---------------|----------|----------|
| | <i>M</i> | (<i>SD</i>) | range | <i>n</i> | <i>M</i> | (<i>SD</i>) | Range | <i>n</i> |
| Age 2 | | | | | | | | |
| DQ | 55.00 | (8.76) | 49 – 71 | 6 | 65.00 | (15.79) | 52 – 93 | 6 |
| ADOS severity | 6.17 | (1.33) | 4 – 8 | 6 | 2.00 | (1.41) | 1 – 4 | 5 |
| Age 4 | | | | | | | | |
| RL | 31.00 | (8.63) | 24 – 42 | 6 | 35.50 | (8.98) | 24 – 48 | 6 |
| EL | 31.00 | (11.01) | 24 – 48 | 6 | 36.50 | (13.07) | 24 – 54 | 6 |
| DQ | 65.33 | (28.21) | 49 – 119 | 6 | 88.50 | (25.07) | 54 – 127 | 6 |
| ADOS severity | 6.67 | (1.37) | 5 – 9 | 6 | 2.00 | (0.89) | 1 – 3 | 6 |
| Difference scores | | | | | | | | |
| DQ | 10.33 | (24.48) | -4 – 60 | 6 | 23.50 | (13.00) | 1 – 38 | 6 |
| ADOS severity | 0.50 | (1.52) | -1 – 3 | 6 | -0.20 | (1.92) | -3 – 2 | 5 |

Note. DQ = developmental quotient, ADOS severity = symptom severity score measured with the Autism Diagnostic Observation Schedule module 1 or 2, RL = receptive language age equivalent, EL = expressive language age equivalent.

Figure Captions

Fig. 1 Overall design of the study

^a The N-CDIs and ESAT were filled out for all children. According to the age of the child, parents filled out the FYI if their child was between 11 and 13 months old, the M-CHAT if their child was between 18 and 24 months old, and the SCQ if their child was older than 24 months or if the parents already filled out the M-CHAT as part of a validation study of the Dutch translation of the M-CHAT.